

## Education Background

Sun Yat-sen University (A Top 10 University in Mainland China)

09/2021-06/2025

**Program:** Bachelor of Engineering in Software Engineering

**GPA:** 3.9/4.0 (Average Score: 91/100)

**Relevant courses:** Principles of Compiler (97), Discrete Mathematics (96), Program Analysis (95), Java and Object-Oriented Design (94), Data Structures and Algorithms (91), Software Analysis and Design (91), etc.

**Awards:** Baosteel Outstanding Student Scholarship (top 1, 2023-2024); 1st Class Scholarship for Outstanding Student (top 5%, 2023-2024); Kingdomcares Scholarship (top 5, 2022-2023); 1st Class Scholarship for Outstanding Student (top 5%, 2022-2023); 3rd Class Scholarship for Outstanding Student (top 30%, 2021-2022); Mathematical Contest in Modeling (MCM), Meritorious Winner (2024); Sun Yat-sen University Programming Design Novice Competition, Second Prize (04/2023).

**International academic experience:** International Asian Studies Program, Chinese University of Hong Kong

## Publication

1. **Jing Gong**, Yanghui Wu, Linxi Liang, Zibin Zheng, Yanlin Wang. “[CoSQA+: Enhancing Code Search Dataset with Matching Code.](#)” *Preprint available at arXiv:2406.11589 [cs.SE]*.
2. **Jing Gong**. “[A Study of Tennis Momentum Based on K-means++ and LightGBM Models.](#)” In *Proceedings of the 2024 2nd International Conference on Mechatronics, IoT, and Industrial Informatics (ICMIII 2024)*, pp. 298-303, Melbourne, VIC, Australia, June 12-14, 2024. Institute of Electrical and Electronics Engineers (IEEE). ISBN: 9798350386639.
3. Chenlin Wu, Xiaoyu He, Zike Li, **Jing Gong**, Zibin Zheng. “[A Historical Trajectory Assisted Optimization Method for Zeroth-Order Federated Learning](#)” *Preprint available at arXiv:2409.15955 [cs.SE]*.

## Research Experience

Repository Level Code Generation, Purdue University

10/2024-Present

*Remote intern, supervised by Prof. Lin Tan*

- Conducted a comprehensive review of existing approaches and benchmarks.
- Formulated innovative strategies to enhance code generation techniques and designed preliminary toy experiments to validate the feasibility and effectiveness of these concepts.
- Developed an agent-based Retrieval-Augmented Generation (RAG) code generation framework to intelligently select retrieval depth and generate code for unknown APIs as supplementary context.

REU: LLM-Orchestrated Bayesian Optimization for AutoML, University of Notre Dame

04/2024-Present

*Project leader, supervised by Prof. Xiangliang Zhang*

- Led a comprehensive study on Bayesian optimization parameter selection, demonstrating how large language models (LLMs) can dynamically select appropriate components based on optimization iteration history and expert knowledge.
- Utilized greedy algorithms and LLM analysis to automatically construct a high-quality dataset for Supervised Fine-Tuning (SFT).
- Fine-tuned the open-source large language model Vicuna, achieving state-of-the-art (SOTA) performance that surpasses existing Bayesian optimization tools on the Bayesmark benchmark.

**SYSU Innovation Project: Code Search Optimization Techniques Base on Large Model Query Optimization and Code Optimization**

02/2024-09/2024

*Project leader, supervised by Prof. Yanlin Wang*

- Completed a prestudy on the optimization of datasets based on large models and researched the cause of model misselection in the code search, as well as existing errors in the dataset through error analysis.

- Constructed the CoSQA+ benchmark by pairing high quality queries with high-quality multiple codes.
- Enhanced the automatic construction process through candidate pair construction by multiple models, automated model annotation and code generation for unmatched queries.
- Introduced a new metric MMRR for assessing multi-code selection.

#### **SYSU Innovation Project: Gradient-free Federated Learning**

**12/2022-12/2023**

*Key member, supervised by Prof. Xiaoyu He*

- Explored gradient-free federated Learning for devices with constrained computational resources; Streamlined federated learning deployment with an adaptive automated parameter selector.
- Developed a scalable federated learning framework that involves randomly selecting clients, placing the model on the server-side, training the clients on the client-side, and uploading the training results.
- Created a dynamic auto-tuning algorithm to optimize execution speed and learning rates.
- Benchmarked optimization methods, such as gradient descent and minibatch-SGD, on datasets like mmist and rcv1 to evaluate their efficiency in logistic regression contexts.

### **Project Experience**

---

#### **2024 Mathematical Contest in Modeling**

**02/2024-02/2024**

*Project Leader*

- Awarded Meritorious Winner.
- Utilized LightGBM to predict game outcomes with over 80% accuracy. Quantified momentum's impact using exponentially weighted moving leverage and K-Means++ clustering.

#### **SYSU Club Management and Recruitment System**

**03/2022-03/2024**

*Project Leader*

- Constructed the overall code framework for the front and back ends using a full-stack development approach.
- Collected requirements and feedback to optimize user experience and interface design.
- Led a cross-functional team through the development lifecycle from ideation to deployment.
- Facilitated efficient communication between stakeholders to gather requirements and feedback.

#### **SSE\_MARKET for the School of Software Engineering**

**03/2022-03/2024**

*Project Leader*

- Developed the back-end framework using Go and the front-end framework using Vue to implement various features.
- Liaised with faculty and student to collect demands and requirements, prioritized user feedback, and improved the platform's features.
- Used by over two-thirds of the college's students and faculty, making it the largest communication center in the college.

#### **The development of the tool: TitleGPT**

**09/2023-01/2024**

*Core member*

- Combined ChatGPT with a custom-built feedforward neural network to create highly engaging headlines.
- Results: the attractiveness of titles generated by TitleGPT increased by 206.06% compared to the original titles, and increased by 158.97% compared to titles directly generated by ChatGPT.

### **Skills & Others**

---

**Language Skills:** English (proficient); Mandarin (native speaker); Cantonese (native speaker)

**Computer Skills:** Proficient: Python/C++/Java; Intermediate: Go/Javascript; Basic: MATLAB

**Competitions:** Participated in 3 mathematical modeling competitions using Python, Numpy/Scipy, PyTorch, and TensorFlow, and respectively studied the computation of Green GDP, the optimization of the distribution of fixed-sun mirror fields, and the effect of momentum on the game of tennis.

**Teaching Assistant:** Software Engineering Training (Intermediate)